

WTF (What Type of Feature)?

Classifying Features for Advanced Data Linking, Searching and Analysis Capabilities

Authors:

Laura Kostanski, Pier-Giorgio Zaccheddu, Linda Merrin, Rob Atkinson.

Conference Sub-Theme:

T6: SDI, Standards, Ontologies, Integration

or

T32: Toponyms on Maps and on GeoDB

Abstract:

Most data about human and natural processes is geographically referenced, usually by some form of identifier such as a town name, administrative area code, address etc. Interpreting these references is the key to integrating and comparing information to support decisions made by a range of government and non-government organisations. The authors contend that understanding the context, or source, of such references is necessary to reliably interpret them, often to distinguish between similarly named but distinctly different ones. For example, a local government area, a suburb name, a tourist destination and a political electorate may have the same name but different semantics and boundaries. Large points of interest databases used for online mapping tools may include many objects with similar names. To understand and disambiguate the identifiers, it is necessary to specify what type of feature is being referred to.

Research into the domains of linked data and the semantic web indicates that reliable, disambiguated geospatial information is fundamentally important to providing clear and consistent understanding of the relationships between many forms of data, including documents, geospatial data and 'big data' interpreted through spatial statistics. Linked data mechanisms for geospatial information, such as that proposed by the CSIRO Spatial Identifier Reference Framework (SIRF), specifies that in the process of integrating data from heterogeneous systems, different feature type classifications increase the complexity of any interpretation or searching process. This leads to a range of issues around ambiguity implied by the use of multiple languages, specific domains of use for Spatial Identifier Reference Datasets (SIRDs) and how finely feature types are differentiated within complex geospatial systems.

One challenge in interpreting identifiers is the localised, domain-relevant or culturally-specific feature type designations used to classify features in each source dataset. The other challenge is locating the source datasets for the features of interest. We define the concept of SIRDs as indexes which list names or codes (such as asset numbers or postcodes) of geographic locations representing a wide range of physical or administrative feature types. Feature type catalogues associated with SIRDs are intended to provide users with tools for clarifying the information associated with data instances. Feature type definitions are also often necessary to formulate queries to select relevant features from these datasets.